

DISK ARRAY CONTROLLER

Publication number: JP2000099281

Publication date: 2000-04-07

Inventor: FUJIMOTO KAZUHISA; TANAKA ATSUSHI;
FUJIBAYASHI AKIRA; KANAI HIROKI; MINOWA
NOBUYUKI

Applicant: HITACHI LTD

Classification:

- **International:** G06F3/06; G06F13/12; G06F3/06; G06F13/12; (IPC1-
7): G06F3/06; G06F3/06

- **European:** G06F3/06D; G06F13/12L

Application number: JP19980264286 19980918

Priority number(s): JP19980264286 19980918

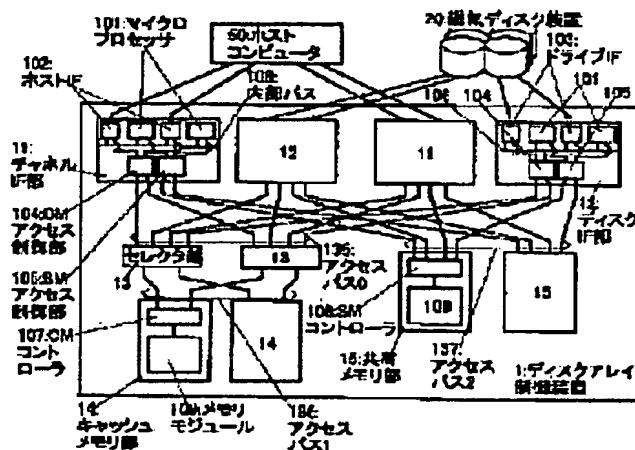
Also published as:

- EP0987623 (A2)
- US6385681 (B1)
- EP0987623 (A3)

[Report a data error here](#)

Abstract of JP2000099281

PROBLEM TO BE SOLVED: To improve throughput by making the style of a connection between plural channel interface and disk interface parts and a cache memory part different from the style of a connection with a shared memory part. **SOLUTION:** A channel interface part 11, a disk interface part 12 and a cache memory part 14 are connected through a selector part 13. Besides, the channel interface part 11, disk interface part 12 and shared memory part 15 are directly connected without interposing the selector part 13. Thus the number of access paths to be directly connected to the cache memory part is reduced, throughput is improved, processing overhead at the selector part is eliminated and time for access to the shared memory part can be shortened.



Data supplied from the esp@cenet database - Worldwide

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-99281

(P2000-99281A)

(43)公開日 平成12年4月7日(2000.4.7)

(51)Int.Cl.⁷
G 0 6 F 3/06識別記号
3 0 5
5 4 0F I
G 0 6 F 3/06テーマコード*(参考)
3 0 5 C 5 B 0 6 5
5 4 0

審査請求 未請求 請求項の数12 O.L (全 19 頁)

(21)出願番号 特願平10-264286

(22)出願日 平成10年9月18日(1998.9.18)

(71)出願人 000005108
 株式会社日立製作所
 東京都千代田区神田駿河台四丁目6番地

(72)発明者 藤本 和久
 東京都国分寺市東恋ヶ窪一丁目280番地
 株式会社日立製作所中央研究所内

(72)発明者 田中 淳
 東京都国分寺市東恋ヶ窪一丁目280番地
 株式会社日立製作所中央研究所内

(74)代理人 100068504
 弁理士 小川 勝男

最終頁に続く

(54)【発明の名称】ディスクアレイ制御装置

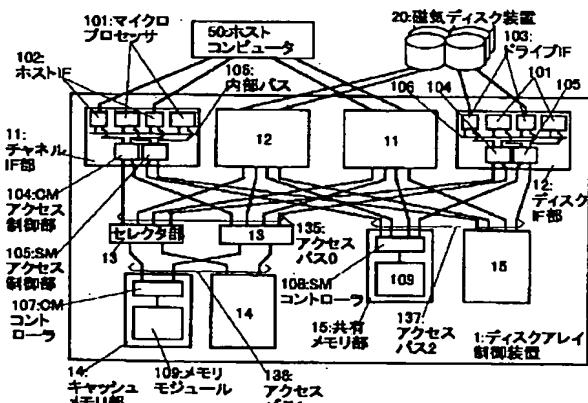
(57)【要約】

【課題】 キャッシュメモリ及び共有メモリに格納されるデータの特性及びこれらのメモリへのアクセス特性を考慮した、スループットが高く、かつ、応答時間の短いディスクアレイ制御装置を提供することにある。

【解決手段】 上記課題は、複数のチャネルIF部と、複数のディスクIF部と、キャッシュメモリ部と、共有メモリ部とを有し、前記複数のチャネルIF部及び前記複数のディスクIF部と前記キャッシュメモリ部との間の接続形式が、前記複数のチャネルIF部及び前記複数のディスクIF部と前記共有メモリ部との間の接続形式と異なることを特徴とするディスクアレイ制御装置により達成される。

【効果】 キャッシュメモリ及び共有メモリへのアクセスバスを増やしスループットを高くする一方、共有メモリへのアクセス時間を短くすることができる。

図1



(2) 開2000-99281 (P2000-9*A)

【特許請求の範囲】

【請求項1】ホストコンピュータとのインターフェースを有する複数のチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、チャネルインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する共有メモリ部とを有し、各チャネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、各ディスクインターフェース部は、前記磁気ディスク装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行するディスクアレイ制御装置において、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間の接続形式が、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間の接続形式と異なることを特徴とするディスクアレイ制御装置。

【請求項2】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続され、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間は、それぞれセレクタ部を介せず直接接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項3】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続され、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ1対1接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項4】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続され、前記複数のチャネルインターフェース部、前記複数のディスクインターフェース部、及び前記共有メモリ部は共有バスに接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項5】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はスイッチを用いた相互結合網によって接続され、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ直接接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項6】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はスイッチを用いた相互結合網によって接続され、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ1対1接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項7】ホストコンピュータとのインターフェースを有する複数のチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、前記複数のチャネルインターフェース部と前記複数のディスクインターフェース部とに接続され、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、前記複数のチャネルインターフェース部と前記複数のディスクインターフェース部とに接続され、チャネルインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する共有メモリ部とを有する制御装置において、

前記キャッシュメモリ部に接続されるアクセスバスの本数は、前記共有メモリ部に接続されるアクセスバスの本数より少ないと特徴とするディスクアレイ制御装置。

【請求項8】セレクタ部をさらに有し、各チャネルインターフェース部及び各ディスクインターフェース部と前記セレクタ部とは、それぞれアクセスバスにより1対1に接続され、前記セレクタ部と前記キャッシュメモリ部とは、アクセスバスにより接続され、各チャネルインターフェース部及び各ディスクインターフェース部と前記セレクタ部と接続するアクセスバスの総数は、前記セレクタ部と前記キャッシュメモリ部とを接続するアクセスバスの総数より多く、

各チャネルインターフェース部及び各ディスクインターフェース部と前記共有メモリ部との間はそれぞれアクセスバスにより1対1に接続されていることを特徴とする請求項7に記載のディスクアレイ制御装置。

【請求項9】ホストコンピュータとの複数のインターフェース部と、磁気ディスク装置との複数のインターフェース部と、磁気ディスク装置のデータを一時的に格納する物理的に独立したキャッシュメモリ部と、制御情報を格納する物理的に独立した共有メモリ部とを有し、前記ホストコンピュータとの各インターフェース部及び前記磁気ディスク装置との各インターフェース部は、それぞれマイクロプロセッサと、前記キャッシュメモリ部へのアクセスを制御するキャッシュメモリ制御部と、前記共有メモリ部へのアクセスを制御する共有メモリアクセス制御部とを有するディスクアレイ制御装置において、各キャッシュメモリアクセス制御部と前記キャッシュメモリ部との間は、セレクタ部を介してアクセスバスによ

(3) 開2000-99281 (P2000-9?A)

り接続されており、各共有メモリアクセス制御部と前記共有メモリ部との間は、それぞれアクセスバスにより直接接続されていることを特徴とするディスクアレイ制御装置。

【請求項10】ホストコンピュータとの複数のインターフェース部と、磁気ディスク装置との複数のインターフェース部と、磁気ディスク装置のデータを一時的に格納する物理的に独立したキャッシュメモリ部と、制御情報を格納する物理的に独立した共有メモリ部とを有し、前記ホストコンピュータとの各インターフェース部及び前記磁気ディスク装置との各インターフェース部は、それぞれマイクロプロセッサと、前記キャッシュメモリ部へのアクセスを制御するキャッシュメモリ制御部と、前記共有メモリ部へのアクセスを制御する共有メモリアクセス制御部とを有するディスクアレイ制御装置において、各キャッシュメモリアクセス制御部と前記キャッシュメモリ部との間は、スイッチを用いた相互結合網によって接続されており、各共有メモリアクセス制御部と前記共有メモリ部との間は、それぞれアクセスバスにより直接接続されていることを特徴とするディスクアレイ制御装置。

【請求項11】各キャッシュメモリ制御部と前記キャッシュメモリ部との間のアクセスバスの帯域幅を、前記ホストコンピュータと前記ホストコンピュータとのインターフェース部との間の最大の全帯域幅の2倍以上としたことを特徴とする請求項9または請求項10の何れかに記載のディスクアレイ制御装置。

【請求項12】前記共有メモリ部及び前記キャッシュメモリ部を複数有し、該複数の共有メモリ部及び前記複数のキャッシュメモリで二重化していることを特徴とする請求項1乃至請求項11の何れかに記載のディスクアレイ制御装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、データを複数の磁気ディスク装置に格納するディスクアレイ装置の制御装置に関する。

【0002】

【従来の技術】半導体記憶装置を記憶媒体とするコンピュータの主記憶のI/O性能に比べて、磁気ディスクを記憶媒体とするディスクサブシステム（以下「サブシステム」という。）のI/O性能は3～4桁程度小さく、従来からこの差を縮めること、すなわちサブシステムのI/O性能を向上させる努力がなされている。サブシステムのI/O性能を向上させるための1つの方法として、複数の磁気ディスク装置でサブシステムを構成し、データを複数の磁気ディスク装置に格納する、いわゆるディスクアレイと呼ばれるシステムが知られている。

【0003】図2は、従来のディスクアレイの構成を示す。ホストコンピュータ50とディスクアレイ制御装置

2との間のデータ転送を実行する複数のチャネルIF部11と、磁気ディスク装置20とディスクアレイ制御装置2間のデータ転送を実行する複数のディスクIF部12と、磁気ディスク装置20のデータを一時的に格納するキャッシュメモリ部14と、ディスクアレイ制御装置2に関する制御情報（例えば、チャネルIF部及びディスクIF部とキャッシュメモリ部14との間のデータ転送制御に関する情報）を格納する共有メモリ部15とを備え、キャッシュメモリ部14および共有メモリ部15は全てのチャネルIF部11及びディスクIF部12からアクセス可能な構成となっている。このディスクアレイでは、チャネルIF部11及びディスクIF部12と共有メモリ部15との間、及び、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間は1対1に接続される。以下、このような接続形態をスター接続と呼ぶ。

【0004】チャネルIF部11は、ホストコンピュータ50と接続するためのインターフェース及びホストコンピュータ50に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスクIF部12は、磁気ディスク装置20と接続するためのインターフェース及び磁気ディスク装置20に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスクIF部12部は、RAID機能の実行も行う。

【0005】図3は、他の従来のディスクアレイの構成を示す。ホストコンピュータ50とディスクアレイ制御装置3間のデータ転送を実行する複数のチャネルIF部11と、磁気ディスク装置20とディスクアレイ制御装置3間のデータ転送を実行する複数のディスクIF部12と、磁気ディスク装置20のデータを一時的に格納するキャッシュメモリ部14と、ディスクアレイ制御装置3に関する制御情報（例えば、チャネルIF部及びディスクIF部とキャッシュメモリ部14との間のデータ転送制御に関する情報）を格納する共有メモリ部15を備え、各チャネルIF部11及びディスクIF部12と共有メモリ部15間は共有バス130で接続され、各チャネルIF部11部及びディスクIF部12とキャッシュメモリ部14間は共有バス131で接続される。以下、このような接続形態を共有バス接続形式と呼ぶ。

【0006】

【発明が解決しようとする課題】ディスクアレイのアーキテクチャーをスケーラブルなものとするには、ディスク制御装置に接続するディスク容量（論理ボリューム数）に応じ、ディスクIF部を増設し、また、必要なホストコンピュータとのチャネル数に応じて、ディスクアレイ制御装置内のチャネルIF部を増設する必要がある。しかし、第3図に示した共有バス接続形式のディスクアレイ制御装置では、一旦実装した共有バスの転送能力をチャネルIF部、ディスクIF部の増設に応じて変

(4) 開2000-99281 (P2000-94) 繼

更することはできないので、チャネルIF部、ディスクIF部の増設に柔軟に対応することが困難である。

【0007】また、第3図に示した共有バス接続形式のディスクアレイ制御装置では、ホストコンピュータとディスクアレイ制御装置との間のデータ転送を実行するチャネルIF部に設けられたホストコンピュータに対する入出力を制御するマイクロプロセッサ、及び磁気ディスク装置とディスクアレイ制御装置との間のデータ転送を実行するディスクIF部に設けられた磁気ディスク装置に対する入出力を制御するマイクロプロセッサに高性能なプロセッサを使用した場合に、これらのプロセッサの性能に比べて、共有バスの転送能力がボトルネックになり、プロセッサの高速化に追従することが困難となる。

【0008】さらに、第3図に示した共有バス接続形式では、共有バスに接続された複数のチャネルIF部（または複数のディスクIF部）の何れかのチャネルIF部（またはディスクIF部）に障害が発生した場合に、障害の発生したチャネルIF部（またはディスクIF部）を特定することが困難である。

【0009】一方、第2図に示したスター接続形式のディスクアレイ制御装置では、共有メモリ部またはキャッシュメモリ部に接続したアクセスバス数に比例して内部バス性能が増加させることができるので、チャネルIF部、ディスクIF部の増設、または使用するプロセッサの性能に応じて、内部バス性能を増加させることができるので、チャネルIF部、ディスクIF部の増設、または使用するプロセッサの性能に応じて、内部バス性能を増加させることができるので、チャネルIF部及びディスクIF部とキャッシュメモリ部との間、チャネルIF部及びディスクIF部と共有メモリ部との間がスター接続されているため、障害の発生したチャネルIF部（またはディスクIF部）を特定することも容易である。

【0010】スター接続形式のディスクアレイ制御装置では、搭載されるチャネルIF部またはディスクIF部の数を増やした場合、チャネルIF部及びディスクIF部とキャッシュメモリ部との間、及びチャネルIF部及びディスクIF部と共有メモリとの間のアクセスバス数も増えることになる。また、ホストコンピュータとディスクアレイ制御装置との間の接続にファイバチャネル等の高速チャネルの採用等により、ディスクアレイ制御装置に要求されるスループットはさらに増大する方向にあり、このスループットの向上の要求を満たすためには、チャネルIF部及びディスクIF部とキャッシュメモリ部との間、及びチャネルIF部及びディスクIF部と共有メモリとの間のアクセスバス数を増やし、内部バス性能を向上させることが必要となる。

【0011】しかし、キャッシュメモリに格納される1つのデータのデータ量は、共有メモリに格納される1つの制御情報のデータ量よりもかなり大きい。一例を挙げれば、メインフレームに接続されるディスク制御装置では、キャッシュメモリに格納される1つのデータは数Kバイト程度（例えば2Kバイト）であるのに対し、共有

メモリに格納される1つの制御情報は数バイト程度（例えば4バイト）である。また、オープン系のホストコンピュータに接続されるディスク制御装置では、キャッシュメモリに格納される1つのデータは数十バイト程度（例えば64バイト）であるのに対し、共有メモリに格納される1つの制御情報は数バイト程度（例えば4バイト）である。したがって、チャネルIF部及びディスクIF部とキャッシュメモリ部との間で転送されるデータ量は、チャネルIF部及びディスクIF部と共有メモリ部との間で転送されるデータ量に比べ、かなり多いので、チャネルIF部及びディスクIF部とキャッシュメモリ部との間のアクセスバスのデータ幅は、チャネルIF部及びディスクIF部と共有メモリ部との間のアクセスバスのデータ幅より広くとする必要がある。例えば、前者のアクセスバスは、16ビット幅のバスで構成され、後者は、4ビット幅のバスで構成される。そのため、チャネルIF部及びディスクIF部とキャッシュメモリ部との間のアクセスバスの本数を増やすと、それらのアクセスバスを接続するキャッシュメモリ部のLSIのピン数が不足するという問題が生じる。また、ディスクアレイ制御装置のホストコンピュータへの応答時間を短くするためにには、共有メモリ部に格納された制御情報へのアクセス時間をできるだけ短くすることも必要である。

【0012】そこで、本発明の目的は、キャッシュメモリ及び共有メモリに格納されるデータの特性及びこれらのメモリへのアクセス特性を考慮した、スループットの高く、ディスクアレイ制御装置、及びそれを用いたサブシステムを提供することにある。

【0013】より具体的には、本発明の目的は、チャネルIF部及びディスクIF部とキャッシュメモリ部との間のアクセスバスはスループットが高く、チャネルIF部及びディスクIF部と共有メモリとの間のアクセスバスはスループットが高く、かつアクセス時間が短いディスクアレイ制御装置、及びそれを用いたサブシステムを提供することにある。

【0014】

【課題を解決するための手段】上記目的は、ホストコンピュータとのインターフェースを有する複数のチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、チャネルインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する共有メモリ部とを有し、各チャネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、各ディスクインターフェース部は、前記磁気ディスク装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行するディスク

(5) 開2000-99281 (P2000-9\$<略縦)

アレイ制御装置において、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間の接続形式が、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間の接続形式と異なることを特徴とするディスクアレイ制御装置により達成される。

【0015】好ましくは、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続し、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間は、それぞれセレクタ部を介せず直接接続する。

【0016】また、好ましくは、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続し、前記複数のチャネルインターフェース部、前記複数のディスクインターフェース部、及び前記共有メモリ部は共有バスに接続する。

【0017】また、好ましくは、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はスイッチを用いた相互結合網によって接続し、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ直接接続する。

【0018】その他、本願が開示する課題、及びその解決方法は、発明の実施形態の欄及び図面により明らかにされる。

【0019】

【発明の実施の形態】以下、本発明の実施例を図面を用いて説明する。

【0020】[実施例1] 図1に、本発明の一実施例を示す。

【0021】ディスクアレイ制御装置1は、ホストコンピュータ50との2つのインターフェース部(チャネルIF部)11と、磁気ディスク装置20との2つのインターフェース部(ディスクIF部)12と、2つのセレクタ部13と、2つのキャッシュメモリ部14と、2つの共有メモリ部15と、アクセスバス0:135と、アクセスバス1:136と、アクセスバス2:137とを有する。

【0022】チャネルIF部11は、ホストコンピュータ50との2つのIF(ホストIF)102と、ホストコンピュータ50に対する入出力を制御する2つのマイクロプロセッサ101と、キャッシュメモリ部14へのアクセスを制御するアクセス制御部(CMアクセス制御部)104と、共有メモリ部15へのアクセスを制御するアクセス制御部(SMアクセス制御部)105とを有

し、ホストコンピュータ50とキャッシュメモリ部14間のデータ転送、及びマイクロプロセッサ101と共有メモリ部15間の制御情報の転送を実行する。マイクロプロセッサ101及びホストIF102は内部バス106によって接続され、CMアクセス制御部104は2つのホストIF102に直接接続されている。また、SMアクセス制御部105は2つのマイクロプロセッサ101に直接接続されている。

【0023】ディスクIF部12は、磁気ディスク装置20との2つのIF(ドライブIF)103と、磁気ディスク装置20に対する入出力を制御する2つのマイクロプロセッサ101と、キャッシュメモリ部14への1つのアクセス制御部(CMアクセス制御部)104と、共有メモリ部15への1つのアクセス制御部(SMアクセス制御部)105を有し、磁気ディスク装置20とキャッシュメモリ部14間のデータ転送、及びマイクロプロセッサ101と共有メモリ部15間の制御情報の転送を実行する。マイクロプロセッサ101及びドライブIF103は内部バス106によって接続され、CMアクセス制御部104は2つのドライブIF103に直接接続されている。また、SMアクセス制御部105は2つのマイクロプロセッサ101に直接接続されている。ディスクIF部はRAID機能の実行も行う。

【0024】キャッシュメモリ部14は、キャッシュメモリ(CM)コントローラ107とメモリモジュール109を有し、磁気ディスク装置20へ記録するデータを一時的に格納する。

【0025】共有メモリ部15は、共有メモリ(SM)コントローラ108とメモリモジュール109とを有し、ディスクアレイ制御装置1の制御情報(例えば、チャネルIF部及びディスクIF部とキャッシュメモリ部14との間のデータ転送制御に関する情報)等を格納する。

【0026】CMアクセス制御部104には2本のアクセスバス0:135を接続し、それらを2つの異なるセレクタ部13にそれぞれ接続する。セレクタ部13には2本のアクセスバス1:136を接続し、それらを2つの異なるCMコントローラ107にそれぞれ接続する。したがってCMコントローラ107には、2つのセレクタ部から1本ずつ、計2本のアクセスバス1:136が接続される。こうすることにより、1つのCMアクセス制御部104から1つのCMコントローラ107へのアクセスルートが2つとなる。これにより、1つのアクセスバスまたはセレクタ部13に障害が発生した場合でも、もう1つのアクセスルートによりキャッシュメモリ部14へアクセスすることが可能となるため、耐障害性を向上させることができる。

【0027】SMアクセス制御部105には2本のアクセスバス2:137を接続し、そのアクセスバス2:137を2つの異なるSMコントローラにそれぞれ接続す

(6) 開2000-99281 (P2000-9A)

る。したがって、SMコントローラ108には、2つのチャネルIF部11及び2つのディスクIF部12から1本ずつ、計4本のアクセスバス2:137が接続される。本実施例では、1つのSMアクセス制御部105と1つのSMコントローラ108の間には1本のアクセスバス2:137を接続したが、このアクセスバス2:137を2本に増やすことにより、1つのSMアクセス制御部105から1つのSMコントローラ108へのアクセスルートが2つとなるため、耐障害性を向上させることができる。

【0028】セレクタ部13には、2つのチャネルIF部11と、2つのディスクIF部12からそれぞれ1本ずつ、計4本のアクセスバス0:135が接続される。また、セレクタ部13には、2つのキャッシュメモリ部14へのアクセスバス1:136が1本ずつ、計2本接続される。

【0029】アクセスバス0:135とアクセスバス1:136の間に上記のようなバス数の関係があるため、セレクタ部13ではチャネルIF部11及びディスクIF部12からの4本のアクセスバス0:135からの要求の内、キャッシュメモリ部14へのアクセスバス1:136の数に相当する2個だけを選択して実行する機能を持つ。

【0030】本実施例の大きな特徴は、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間の接続形式と、チャネルIF部11及びディスクIF部12と共有メモリ部15との間の接続形式とが異なる点にある。このような構成にした理由を図1及び図2を用いて説明する。図2に示したスター接続形式のディスクアレイ制御装置2では、ディスクアレイ制御装置2に搭載されるチャネルIF部11またはディスクIF部12の数を増やした場合、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間、及びチャネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバス数も増えることになる。また、スター接続形式のディスクアレイ制御装置2において、スループットを向上させるためには、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間、及びチャネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバス数を増やし、内部バス性能を向上させることができある。

【0031】しかし、キャッシュメモリ部14に格納される1つのデータのデータ量は、共有メモリ部15に格納される1つの制御情報のデータ量よりもかなり大きい。一例を挙げれば、メインフレームに接続されるディスク制御装置では、キャッシュメモリ部14に格納される1つのデータは数Kバイト程度（例えば2Kバイト）であるのに対し、共有メモリ部15に格納される1つの制御情報は数バイト程度（例えば4バイト）である。ま

た、オープン系のホストコンピュータに接続されるディスク制御装置では、キャッシュメモリ部14に格納される1つのデータは数十バイト程度（例えば64バイト）であるのに対し、共有メモリ部15に格納される1つの制御情報は数バイト程度（例えば4バイト）である。したがって、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間で転送されるデータ量は、チャネルIF部11及びディスクIF部12と共有メモリ部15との間で転送されるデータ量に比べかなり多いので、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間のアクセスバスのデータ幅は、チャネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバスのデータ幅より広くとる必要がある。例えば、前者のアクセスバスは16ビット幅のバスで構成され、後者は、8ビット幅のバスで構成される。そのため、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間のアクセスバス本数を増やすと、キャッシュメモリ部14内のキャッシュメモリコントローラ（図2ではキャッシュメモリコントローラを図示していない）のLSIのピン数不足、またはキャッシュメモリ部14を実装するパッケージにおいてコネクタのピン数不足という問題が生じる。そこで、本実施例では、図1に示すように、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間をセレクタ部13を介して接続することにより、キャッシュメモリ部14に直接接続されるアクセスバス数を削減している。

【0032】一方、上述したように、共有メモリ部13へ格納する1つの制御情報のデータ長はキャッシュメモリ部14に格納する1つのデータのデータ長に比べかなり小さいので、チャネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバスのデータ幅は、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間のアクセスバスのデータ幅の半分以下とすることが可能である。したがって、共有メモリ部15へのアクセスバス数を増やしても共有メモリ部内の共有メモリコントローラ（図2では共有メモリコントローラを図示していない）のLSIのピン数不足等の問題が生じることは少ない。

【0033】また、ディスクアレイ制御装置1のホストコンピュータ50への応答時間を短くするためには、共有メモリ部15に格納される制御情報へのアクセス時間をできるだけ短くする必要もある。しかし、図1に示したCMアクセス制御部104とCMコントローラ107間のように、SMアクセス制御部105とSMコントローラ108との間をセレクタ部を介して接続すると、セレクタ部での処理のオーバーヘッドにより、共有メモリ部15に格納される制御情報へのアクセス時間を短くすることができない。

【0034】そこで、本実施例では、チャネルIF部1

(7) 開2000-99281 (P2000-9(緯)

1及びディスクIF部12と共有メモリ部15との間をセレクタ部を介さず直接接続することにより、チャネルIF部11及びディスクIF部12と共有メモリ部15間に複数のアクセスバスを設け、スループットを向上させる一方、セレクタ部での処理オーバーヘッドをなくし、チャネルIF部11、及びディスクIF部12から共有メモリ部15へのアクセス時間を短縮している。
【0035】なお、本実施例では、耐障害性の向上という観点から、セレクタ部13、キャッシュメモリ部14、及び共有メモリ部15をそれぞれ二重化しているが、これらを二重化しなくても、上述の効果を得られるることは言うまでもない。

【0036】図4は、CMアクセス制御部104内の構成を示している。CMアクセス制御部104は、セレクタ302と、アドレス、コマンド、データを一時格納するパケットバッファ303と、セレクタ部13に繋がるアクセスバス0:135とのバスIF301と、データのエラーチェック部300と、データ転送制御部310を有する。セレクタ302の2つのポートはデータ線210でホストIF102あるいはドライブIF103に接続される。また、セレクタ302の他の2つのポートはバスIF301に接続される。バスIF301はアクセスバス0:135でセレクタ部13に接続される。データ転送制御部310は、制御線1:211でホストIF102あるいはドライブIF103に接続され、制御線2:212でセレクタ部13内のデータ転送制御部315に接続される。また、データ転送制御部310は、アビタ308によりホストIF102あるいはドライブIF103からのアクセス要求のアビトレーションを行い、セレクタ302の切り替えを行う。

【0037】図6は、セレクタ部13内の構成を示している。セレクタ部13は、チャネルIF部11及びディスクIF部12に繋がるアクセスバス0:135との4つのバスIF301と、CMコントローラ107に繋がるアクセスバス1:136との2つのバスIF301と、両者間を互いに接続するセレクタ306と、パケットバッファ303と、データのエラーチェック部300と、CMアクセス制御部104から送出されたアドレス及びコマンドを解析するアドレス・コマンド(adrl, cmdl)解析部305と、データ転送制御部315を有する。データ転送制御部315は、制御線2:212でCMアクセス制御部104内のデータ転送制御部310に接続され、制御線3:213でCMコントローラ107内のデータ転送制御部315に接続される。また、データ転送制御部315は、アビタ308により、adrl, cmdl解析部305で解析した4つのアクセスバス0:135からのアクセス要求のアビトレーションを行い、セレクタ306の切り替えを行う。パケットバッファ303は、アクセスバス0:135側のバスとアクセスバス1:136側のバスでデータ転送速度に差があ

る場合、速度差を吸収するために、転送するデータの一部または全部をバッファリングする。

【0038】adrl, cmdl解析部305は、アドレス及びコマンドを格納するバッファと、adrl抽出部と、cmdl抽出部を有する(図示していない)。adrl, cmdl解析部305では、CMアクセス制御部104に接続される4本のアクセスバス0:135それぞれに1つずつ割り当てられたバッファに、アドレス、コマンドを格納する。adrl抽出部及びcmdl抽出部では、アクセスするCMコントローラ107とアクセスの種類を割り出し、データ転送制御部315内のアビタ308へ送出する。

【0039】図7は、キャッシュメモリ部14内の構成を示している。キャッシュメモリ部14は、CMコントローラ107とメモリモジュール109を有する。CMコントローラ107は、セレクタ部13に繋がるアクセスバス1:136との2つのバスIF301と、セレクタ304と、データを一時格納するパケットバッファ303と、データのエラーチェック部300と、メモリモジュール109へのアクセスを制御するメモリ制御部307と、CMアクセス制御部104から送出されたアドレス及びコマンドを解析するadrl, cmdl解析部305と、データ転送制御部315を有する。データ転送制御部315は、制御線3:213でセレクタ部13内のデータ転送制御部315に接続される。また、データ転送制御部315は、アビタ308により、adrl, cmdl解析部305で解析した2本のアクセスバス1:136からのアクセス要求のアビトレーションを行い、セレクタ304の切り替えを行う。

【0040】adrl, cmdl解析部305は、バッファと、adrl抽出部と、cmdl抽出部を有する(図示していない)。adrl, cmdl解析部305では、CMコントローラ107に接続される2本のアクセスバス1:136それぞれに1つずつ割り当てられたバッファに、アドレス、コマンドを格納する。adrl抽出部及びcmdl抽出部では、アクセスするメモリのアドレスとアクセスの種類を割り出し、メモリ制御部307へ送出する。また、2本のアクセスバス1:136からのアクセス要求をデータ転送制御部315内のアビタ308へ送出する。

【0041】次に、キャッシュメモリ部14へのアクセス時の手順について述べる。キャッシュメモリ部14へアクセスする場合、マイクロプロセッサ101は、ホストIF102あるいはドライブIF103へ、キャッシュメモリ部14へのアクセス開始を指示する。

【0042】アクセス開始の指示を受けたホストIF102あるいはドライブIF103は、制御線1:211によりCMアクセス制御部104内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線210を通してアドレス、コマンド、デ

(8) 開2000-99281 (P2000-9%" 略縫

ータ（データの書き込み時のみ）を送出する。

【0043】CMアクセス制御部104は、データ線210を通して送られてきたアドレス、コマンド、データ（データの書き込み時のみ）をパケットバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セレクタ302を切り替える。

【0044】図9は、キャッシュメモリ部14へデータを書き込む場合の、CMアクセス制御部104からCMコントローラ107へのアクセスの流れを示している。CMアクセス制御部104内のデータ転送制御部310は、アービトレーションによってアクセスバス0:135の使用権が決定されると、制御線2:212によってセレクタ部13内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ501)。続いて、アドレス及びコマンドを送出する(ステップ502)。

【0045】セレクタ部13内のデータ転送制御部315は、CMアクセス制御部104からREQ信号を受け取ると、次にアクセスバス0:135を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行う(ステップ503)。アービトレーションの結果、アクセスバス1:136への接続権を得たら、データ転送制御部315はセレクタ306を切り替える(ステップ504)とともに、制御線2:212により、CMアクセス制御部104内のデータ転送制御部310へ、アクセスバス1:136への接続権が得られたことを示す信号(ACK)を返す(ステップ505)。次にデータ転送制御部315は、制御線3:213によってCMコントローラ107内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ506)。続いて、アドレス及びコマンドを送出する(ステップ507)。

【0046】CMアクセス制御部104はACK信号を受けると、パケットバッファ303からデータを読み出し、セレクタ302、バスIF301を介してアクセスバス0:135へ送出する。セレクタ部13は、アクセスバス0:135を通して送られてきたデータを、バスIF301及びセレクタ306を介してアクセスバス1:136へ送出する(ステップ509)。

【0047】CMコントローラ107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスバス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行い(ステップ508)、セレクタ304を切り替える。アクセスバス1:136を通して送られてくるデータはパケットバッファ303に格納する。アービトレーションの結果、メモリモジュール10

9へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ510)。次に、パケットバッファ303からデータを読み出し、セレクタ304を介してメモリモジュール109へ書き込む(ステップ511)。

【0048】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ512)。次に、ステータスをセレクタ部13を介してCMアクセス制御部104へ送出する(ステップ513)。セレクタ部13内のデータ転送制御部315はステータスを受け取ると、CMコントローラ107へのREQ信号をオフする(ステップ514)。CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、セレクタ部13へのREQ信号をオフする(ステップ515)。セレクタ部13内のデータ転送制御部315はCMアクセス制御部104からのREQ信号のオフを確認すると、CMアクセス制御部104へのACK信号をオフする(ステップ516)。

【0049】CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、制御線1:211により、ホストIF102あるいはドライブIF103へキャッシュメモリ部14へのアクセスの終了を報告する。

【0050】キャッシュメモリ部14からデータを読み出す場合の、CMアクセス制御部104からCMコントローラ107へのアクセスの流れは、ステップ501から508までとステップ512以降は、データの書き込みの場合と同じである。

【0051】ここでCMアクセス制御部104は、ステップ505でACK信号を受けると、データの受信待ち状態に入る。

【0052】ステップ508でメモリアクセス権を得ると、CMコントローラ107はメモリモジュール109からデータを読み出し、セレクタ304、バスIF301を介してアクセスバス1:136にデータを送出する。

【0053】セレクタ部13は、アクセスバス1:136を通してデータを受信すると、バスIF301及びセレクタ306を介してアクセスバス0:135にデータを送出する。

【0054】CMアクセス制御部104は、アクセスバス0:135を通してデータを受信すると、セレクタ302、データ線210を介してホストIF102あるいはドライブIF103へデータを送出する。

【0055】図5は、SMアクセス制御部105内の構成を示している。SMアクセス制御部104は、セレクタ302と、アドレス、コマンド、データを一時格納するパケットバッファ303と、SMコントローラ108

(9) 開2000-99281 (P2000-9 織)

に繋がるアクセスバス2:137とのバスIF301と、データのエラーチェック部300と、データ転送制御部310を有する。セレクタ302の2つのポートはデータ線220でマイクロプロセッサ101に接続される。また、セレクタ302の他の2つのポートはバスIF301に接続される。バスIF301はアクセスバス2:137でSMコントローラ108に接続される。データ転送制御部310は、制御線5:221でマイクロプロセッサ101に接続され、制御線6:222でSMコントローラ108内のデータ転送制御部315に接続される。また、データ転送制御部310は、アビタ308によりマイクロプロセッサ101からのアクセス要求のアビトレーションを行い、セレクタ302の切り替えを行う。

【0056】図8は、共有メモリ部15内の構成を示している。共有メモリ部15は、SMコントローラ108とメモリモジュール109を有する。SMコントローラ108は、SMアクセス制御部105に繋がるアクセスバス2:137との4つのバスIF301と、セレクタ309と、データを一時格納するパケットバッファ303と、データのエラーチェック部300と、メモリモジュール109へのアクセスを制御するメモリ制御部307と、SMアクセス制御部105から送出されたアドレス及びコマンドを解析するadr、cmd解析部305と、データ転送制御部315を有する。データ転送制御部315は、制御線6:222でSMアクセス制御部105内のデータ転送制御部310に接続される。また、データ転送制御部315は、アビタ308により、adr、cmd解析部305で解析した4本のアクセスバス2:137からのアクセス要求のアビトレーションを行い、セレクタ309の切り替えを行う。

【0057】adr、cmd解析部305は、バッファと、adr抽出部と、cmd抽出部を有する(図示していない)。adr、cmd解析部305では、SMコントローラ108に接続される4本のアクセスバス2:137それぞれに1つずつ割り当てられたバッファに、アドレス、コマンドを格納する。adr抽出部及びcmd抽出部では、アクセスするメモリのアドレスとアクセスの種類を割り出し、メモリ制御部307へ送出する。また、4本のアクセスバス2:137からのアクセス要求をデータ転送制御部315内のアビタ308へ送出する。

【0058】次に、共有メモリ部15へのアクセス時の手順について述べる。共有メモリ部15へアクセスする場合、マイクロプロセッサ101は、制御線5:221によりSMアクセス制御部105内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線220を通してアドレス、コマンド、データ(データの書き込み時のみ)を送出する。

【0059】SMアクセス制御部105は、データ線2

20を通して送られてきたアドレス、コマンド、データ(データの書き込み時のみ)をパケットバッファ303に格納する。データ転送制御部310はアビトレーションを行ってバスIF301の使用権を決定し、セレクタ302を切り替える。

【0060】図10は、共有メモリ部15へデータを書き込む場合の、SMアクセス制御部105からSMコントローラ108へのアクセスの流れを示している。SMアクセス制御部105内のデータ転送制御部310は、アビトレーションによってアクセスバス2:137の使用権が決定されると、制御線6:222によってSMコントローラ108へアクセス開始を示す信号(REQ)を出す(ステップ601)。続いて、アドレス、コマンド、及びデータを連続して送出する(ステップ602)。

【0061】SMコントローラ108内のデータ転送制御部315は、制御線6:222によってREQ信号を受け取ると、次にアクセスバス2:137を通して送られてくるアドレス、コマンド、及びデータを受信する。アドレスとコマンドは、adr、cmd解析部305で解析し、アクセス要求に基づいてアビトレーションを行い(ステップ603)、セレクタ309を切り替える。データはパケットバッファ303に格納する。アビトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ604)。次に、パケットバッファ303からデータを読み出し、セレクタ309を介してメモリモジュール109へ書き込む(ステップ605)。

【0062】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ606)。次に、ステータスをSMアクセス制御部105へ送出する(ステップ607)。SMアクセス制御部105内のデータ転送制御部310はステータスを受け取ると、SMコントローラ108へのREQ信号をオフする(ステップ608)。

【0063】SMアクセス制御部105内のデータ転送制御部310はステータスを受け取ると、制御線5:221により、マイクロプロセッサ101へ共有メモリ部15へのアクセスの終了を報告する。

【0064】共有メモリ部15からデータを読み出す場合のSMアクセス制御部105からSMコントローラ108へのアクセスの流れは、ステップ601から604までとステップ606以降は、データの書き込みの場合と同じである。

【0065】ステップ604でメモリアクセスの前処理を行った後、SMコントローラ108はメモリモジュール109からデータを読み出し、セレクタ309、バス

(10) 月2000-99281 (P2000-9) 織

I F 3 0 1 を介してアクセスバス2：1 3 7 にデータを送出する。

【0066】SMアクセス制御部1 0 5 は、アクセスバス2：1 3 7 を通してデータを受信すると、セレクタ3 0 2 、データ線2 2 0 を介してマイクロプロセッサ1 0 1 へデータを送出する。

【0067】本実施例のディスクアレイ制御装置1 では、ホストコンピュータ5 0 とのチャネルを2つ有するチャネルI F 部1 2 を複数搭載しており、それらのチャネルをそれぞれ異なるホストコンピュータに接続することが可能である。そうした場合、ホストコンピュータに接続した各チャネルからの要求全てを並列に処理する必要がある。

【0068】ところでディスクアレイ制御装置1 では、ホストコンピュータ5 0 へデータを読み出す場合、磁気ディスク装置2 0 に格納されたデータをディスクI F 部1 2 を介してキャッシュメモリ部1 4 へ書き込み、そのデータをキャッシュメモリ部1 4 から読み出して、チャネルI F 部1 1 を介してホストコンピュータ5 0 へ送る。またホストコンピュータ5 0 からディスクアレイ制御装置1 へデータを書き込む場合は、ホストコンピュータ5 0 からチャネルI F 部1 2 へ送られてきたデータをキャッシュメモリ部1 4 へ書き込み、そのデータをキャッシュメモリ部1 4 から読み出して、ディスクI F 部1 2 を介して磁気ディスク装置2 0 に書き込む。さらにデータのバリティを生成して磁気ディスク装置2 0 に書き込むため、ディスクI F 部1 2 とキャッシュメモリ部1 4 間でさらに2～3回のアクセスが行われる。

【0069】したがって、ホストコンピュータに接続した各チャネルからの要求全てを並列に処理するためには、チャネルI F 部1 1 及びディスクI F 部1 2 とキャッシュメモリ部1 4 間のスループットをホストコンピュータ5 0 とチャネルI F 部1 1 間の最大のスループットの2倍以上にしなければならない。

【0070】本実施例では、CMアクセス制御部1 0 4 とCMコントローラ1 0 7 間のアクセスバスの帯域幅、及びCMコントローラ1 0 7 とメモリモジュール1 0 9 間の全帯域幅を、チャネルI F 部1 1 とホストコンピュータ5 0 間の最大の帯域幅の2倍以上に設定する。これにより、全チャネルI F 部1 1 を並列に動作させることが可能となる。

【0071】本実施例によれば、チャネルI F 部1 1 及びディスクI F 部1 2 とキャッシュメモリ部1 4 間のスループットを高くすること、チャネルI F 部1 1 及びディスクI F 部1 2 と共有メモリ部1 5 間のスループットを高く、且つアクセス時間を短くすることの両方が可能となる。これによって、スループットが高く、且つ応答時間の短いディスクアレイ制御装置を提供できる。

【0072】ここで、図1 5 に示すように、チャネルI F 部1 1 及びディスクI F 部1 2 とキャッシュメモリ部

1 4 間をスイッチ(SW)1 6 を用いた相互結合網1 4 0 で接続する。この場合にも、図1 に示したセレクタ部1 3 を介して接続した構成と同様に、キャッシュメモリ部1 4 へ複数のアクセスバスを設けることができるため、スループットを高めることができる。

【0073】また図1 6 に示すように、1つのCMアクセス制御部1 0 4 へ接続されるアクセスバス0：1 3 5 の本数を図1 の構成の倍の4本に増やしたディスクアレイ制御装置1 においても、本実施例を実施する上で問題はない。ホストI F 及びドライブI F として、今後はファイバーチャネル等のスループットが1 0 0 M B / s 以上の高速I F が使用されることが多くなると考えられる。チャネルI F 部1 1 及びディスクI F 部1 2 でのスループットのバランスを考えると、1つのCMアクセス制御部1 0 4 に繋がる全アクセスバス0：1 3 5 のスループットは、チャネルI F 部1 1 内の全ホストI F 1 0 2 、またはディスクI F 部1 2 内の全ドライブI F 1 0 3 のスループットと同等以上にする必要がある。上記のようにファイバーチャネル等の高速I F を使用する場合は、図1 6 に示すように、1つのCMアクセス制御部1 0 4 に繋がるアクセスバス0：1 3 5 の本数を増やすことで、アクセスバス0：1 3 5 のスループットをホストI F 1 0 2 あるいはドライブI F 1 0 3 のスループット以上にすることができる。

【0074】なお、図1 5 及び図1 6 のディスクアレイ制御装置では、耐障害性の向上という観点から、セレクタ部1 3 、キャッシュメモリ部1 4 、及び共有メモリ部1 5 をそれぞれ二重化しているが、これらを二重化しなくとも、上述の効果を得られることは言うまでもない。

【0075】【実施例2】図1 において、2つのキャッシュメモリ部1 4 間で、メモリ領域の全部または、一部を二重化し、キャッシュメモリ部1 4 へのデータの書き込み時に二重化した2つの領域に同じデータを書き込むことにより、データの信頼性を上げることが可能となる。

【0076】二重化した2つのキャッシュメモリ部1 4 へデータを書き込む場合の手順は以下のようになる。

【0077】アクセス開始の指示を受けたホストI F 1 0 2 あるいはドライブI F 1 0 3 は、制御線1：2 1 1 によりCMアクセス制御部1 0 4 内のデータ転送制御部3 1 0 へアクセス開始を示す信号を送出する。それとともに、データ線2 1 0 を通して2つのアドレス、コマンドと、1つのデータを送出する。

【0078】CMアクセス制御部1 0 4 は、データ線2 1 0 を通して送られてきた2つのアドレス、コマンドと、1つのデータをパケットバッファ3 0 3 に格納する。データ転送制御部3 1 0 はアビトレイションを行ってバスI F 3 0 1 の使用権を決定し、セレクタ3 0 2 を切り替える。

【0079】図1 1 は、二重化した2つのキャッシュメ

(11) 月2000-99281 (P2000-9A)

モリ部14へデータを書き込む場合の、CMアクセス制御部104から2つのCMコントローラa、b:107へのアクセスの流れを示している。CMアクセス制御部104内のデータ転送制御部310は、アビトリエーションによってアクセスバス0:135の使用権が決定されると、制御線2:212によってセレクタ部13内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ701)。続いて、アドレス及びコマンドを2つ連続して送出する(ステップ702)。

【0080】セレクタ部13内のデータ転送制御部315は、CMアクセス制御部104からREQ信号を受け取ると、次にアクセスバス0:135を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアビトリエーションを行う(ステップ703)。アビトリエーションの結果、CMコントローラa、b:107への2本のアクセスバス1:136への接続権の両方を得たら、データ転送制御部315はセレクタ306を切り替える(ステップ704)とともに、制御線2:212により、CMアクセス制御部104内のデータ転送制御部310へ、CMコントローラa、b:107への接続権が得られたことを示す信号(ACK)を返す(ステップ705)。次にデータ転送制御部315は、制御線3:213によって2つのCMコントローラ107内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ706a、706b)。続いて、CMコントローラa、b:107へアドレス及びコマンドを1つずつ送出する(ステップ707a、707b)。

【0081】CMアクセス制御部104はACK信号を受けると、パケットバッファ303からデータを読み出し、セレクタ302、バスIF301を介してアクセスバス0:135へ送出する(ステップ709)。セレクタ部13は、アクセスバス0:135を通して送られてきた1つのデータを、バスIF301及びセレクタ306を介して2つのアクセスバス1:136の両方へ送出する(ステップ709a、709b)。

【0082】CMコントローラa、b:107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスバス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアビトリエーションを行い(ステップ708a、708b)、セレクタ304を切り替える。アクセスバス1:136を通して送られてくるデータはパケットバッファ303に格納する。アビトリエーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ710a、710b)。次に、パケットバッファ303からデータを読み

出し、セレクタ304を介してメモリモジュール109へ書き込む(ステップ711a、711b)。

【0083】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ712a、712b)。次に、ステータスをセレクタ部13を介してCMアクセス制御部104へ送出する(ステップ713a、713b)。セレクタ部13内のデータ転送制御部315はステータスを受け取ると、CMコントローラa、b:107へのREQ信号をそれぞれオフする(ステップ714a、714b)。また、セレクタ部13はCMコントローラa、b:107の両方からステータスを受け取ったら、それらを続けてCMアクセス制御部へ送出する(ステップ715)。CMアクセス制御部104内のデータ転送制御部310は2つのステータスを受け取ると、セレクタ部13へのREQ信号をオフする(ステップ715)。セレクタ部13内のデータ転送制御部315はCMアクセス制御部104からのREQ信号のオフを確認すると、CMアクセス制御部104へのACK信号をオフする(ステップ716)。

【0084】CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、制御線1:211により、ホストIF102あるいはドライブIF103へキャッシュメモリ部14へのアクセスの終了を報告する。

【0085】上記のように二重化したキャッシュメモリ部14にデータを二重に書き込む場合、二重に書き込むデータの一方のデータの書き込み処理を行っている間に、もう一方のデータが他のアクセスバスからの書き込み要求によって書き換えられるのを防ぐ必要がある。本実施例では、共有メモリ部15にキャッシュメモリ部14のディレクトリを格納し、キャッシュメモリ部14にアクセスする前に必ず、共有メモリ部15に格納したディレクトリにアクセス中を示すビットを立てる。これにより、キャッシュメモリ部14内の同じアドレスには同時に1つのアクセス要求しか発行されないため、二重に書き込むデータの一方のデータの書き込み処理を行っている間に、もう一方のデータが他のアクセスバスからの書き込み要求によって書き換えられるのを防ぐことができる。

【0086】[実施例3]ディスクアレイ制御装置1では、キャッシュメモリ部14を複数設けた場合、あるキャッシュメモリ部14から別のキャッシュメモリ部14へデータをコピーする機能が要求される。この機能は、以下に述べる手順で実現できる。

【0087】アクセス開始の指示を受けたホストIF102あるいはドライブIF103は、制御線1:211によりCMアクセス制御部104内のデータ転送制御部310へアクセス開始を示す信号を送出する。それと

(12) 月2000-99281 (P2000-9T縦)

もに、データ線210を通して2つのアドレス、コマンドを送出する。2つのうちの1つのアドレス及びコマンドは、コピー元のアドレスとリードコマンドで、もう1つのアドレス及びコマンドはコピー先のアドレスとライトコマンドである。ここでは、CMコントローラa:107をコピー元、CMコントローラbをコピー先として説明する。

【0088】CMアクセス制御部104は、データ線210を通して送られてきた2つのアドレス、コマンドをパケットバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セレクタ302を切り替える。

【0089】図12は、2つのキャッシュメモリ部間でデータをコピーする場合の、CMアクセス制御部104からCMコントローラa、b:107へのアクセスの流れを示している。CMアクセス制御部104内のデータ転送制御部310は、アービトレーションによってアクセスバス0:135の使用権が決定されると、制御線2:212によってセレクタ部13内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ801)。続いて、アドレス及びコマンドを2つ連続して送出する(ステップ802)。

【0090】セレクタ部13内のデータ転送制御部315は、CMアクセス制御部104からREQ信号を受け取ると、次にアクセスバス0:135を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行う(ステップ803)。アービトレーションの結果、CMコントローラa、b:107への2つのアクセスバス1:136への接続権の両方を得たら、データ転送制御部315はセレクタ306を切り替える(ステップ804)とともに、制御線2:212により、CMアクセス制御部104内のデータ転送制御部310へ、CMコントローラa、b:107両方への接続権が得られたことを示す信号(ACK)を返す(ステップ805)。次にデータ転送制御部315は、制御線3:213によってCMコントローラa、b:107内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ806a、806b)。続いて、CMコントローラa、b:107へそれぞれのアドレス及びコマンドを送出する(ステップ807a、807b)。

【0091】CMアクセス制御部104はACK信号を受けると、アクセスの終了を知らせるステータスの受信待ち状態に入る。

【0092】コピー元のCMコントローラa:107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスバス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要

求に基づいてアービトレーションを行い(ステップ808)、セレクタ304を切り替える。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ809)。次に、メモリモジュール109からデータを読み出し(ステップ810)、セレクタ304を介してアクセスバス1:136へ送出する(ステップ811a)。

【0093】セレクタ部13は、アクセスバス1:136を通してCMコントローラa:107から送られてきたデータをCMコントローラbに繋がるアクセスバス1:136へ送出する。(ステップ811b)。

【0094】CMコントローラb:107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスバス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析する。その後、データを受信し始めるのを待って、メモリアクセスのアービトレーションに参加する(ステップ812)。アクセスバス1:136を通して送られてくるデータはパケットバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ813)。次に、パケットバッファ303からデータを読み出し、セレクタ304を介してメモリモジュール109へ書き込む(ステップ814)。

【0095】CMコントローラa、b:107は、それぞれCMメモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ815、818)。次に、ステータスをセレクタ部13へ送出する(ステップ816、819)。

【0096】セレクタ部13内のデータ転送制御部315はステータスを受け取ると、CMコントローラa、b:107へのREQ信号をそれぞれオフする(ステップ817、821)。また、セレクタ部13は2つのCMコントローラa、b:107の両方からステータスを受け取ったら、それらを続けてCMアクセス制御部へ送出する(ステップ820)。CMアクセス制御部104内のデータ転送制御部310は2つのステータスを受け取ると、セレクタ部13へのREQ信号をオフする(ステップ822)。セレクタ部13内のデータ転送制御部315はCMアクセス制御部104からのREQ信号のオフを確認すると、CMアクセス制御部104へのACK信号をオフする(ステップ823)。

【0097】CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、制御線1:2

(13) 月2000-99281 (P2000-9KA)

11により、ホストIF102あるいはドライブIF103へキャッシュメモリ部14へのアクセスの終了を報告する。

【0098】ディスクアレイ制御装置1ではまた、1つのキャッシュメモリ部14内のあるアドレスから別のアドレスへデータをコピーする機能も要求される。

【0099】この機能は、図9で示したデータの書き込み時の手順において、ステップ511のメモリモジュール109へのライトアクセスの代わりに、メモリモジュール109からデータを読み出してCMコントローラ107内のパケットバッファ303に格納し、続けてそのデータをメモリモジュール109へ書き込むという処理を行うことによって実現できる。

【0100】[実施例4]図1において、2つの共有メモリ部15間で、メモリ領域の全部または、一部を二重化し、共有メモリ部15へのデータの書き込み時に二重化した2つの領域に同じデータを書き込むことにより、データの信頼性を上げることが可能となる。

【0101】二重化した2つの共有メモリ部15へデータを書き込む場合の手順は、以下のようになる。

【0102】マイクロプロセッサ101は、制御線5:221によりSMアクセス制御部105内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線220を通して2つのアドレス、コマンドと、1つのデータを送出する。

【0103】SMアクセス制御部105は、データ線220を通して送られてきた2つのアドレス、コマンドと、1つのデータをパケットバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セレクタ302を切り替える。

【0104】図13は、二重化した2つの共有メモリ部15へデータを書き込む場合の、SMアクセス制御部105から2つのSMコントローラ108へのアクセスの流れを示している。2つの共有メモリ部を二重化する場合、一方をマスタ、もう一方をスレーブに設定する。SMアクセス制御部105内のデータ転送制御部310は、アービトレーションによって、まずマスタ側のSMコントローラ108へのアクセスパス2:137の使用権を決定し、制御線6:222によってマスタ側のSMコントローラ108内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ901)。続いて、アドレス、コマンド、及びデータを連続して送出する(ステップ902)。

【0105】マスタ側のSMコントローラ108内のデータ転送制御部315は、制御線6:222によってREQ信号を受け取ると、次にアクセスパス2:137を通して送られてくるアドレス、コマンド、及びデータを受信し、addr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行い(ステップ

903)、セレクタ309を切り替える。データはパケットバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ904)。次に、パケットバッファ303からデータを読み出し、セレクタ309を介してメモリモジュール109へ書き込む(ステップ905)。

【0106】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ906)。次に、ステータスをSMアクセス制御部105へ送出する(ステップ907)。

【0107】SMアクセス制御部105はステータスを受け取ったら、マスタ側のSMコントローラ108へのアクセスパス2:137の使用権を開放せずに、アービトレーションによって、スレーブ側のSMコントローラ108へのアクセスパス2:137の使用権を決定する。その後のスレーブ側のSMコントローラ108へのアクセス手順(ステップ908~914)は、マスタ側のSMコントローラ108へのアクセス手順(ステップ901~907)と同様である。

【0108】SMアクセス制御部105内のデータ転送制御部310はスレーブ側のSMコントローラ108からステータスを受け取ると、マスタ側とスレーブ側両方のSMコントローラ108へのREQ信号をオフする(ステップ715a、715b)。

【0109】二重化した共有メモリ部15にデータを二重に書き込む場合、上記のように共有メモリ部15をマスタとスレーブに分け、マスタ側、スレーブ側の順でデータを書き込み、スレーブ側へのデータの書き込みが終わるまで両方のアクセスパス2:137を開放しない。これによりデータを書き込む順番が保証され、二重に書き込むデータの一方のデータの書き込み処理を行っている間に、もう一方のデータが他のアクセスパスからの書き込み要求によって書き換えられるのを防ぐことができる。

【0110】SMアクセス制御部105内のデータ転送制御部310はステータスを受け取ると、制御線5:221により、マイクロプロセッサ101へ共有メモリ部15へのアクセスの終了を報告する。

【0111】[実施例5]図14に、実施例1のディスクアレイ制御装置1のチャネルIF部11、ディスクIF部12、セレクタ部13、キャッシュメモリ部14、共有メモリ部15を実装するときの構成を示す。

【0112】チャネルIF部11、ディスクIF部12、セレクタ部13、キャッシュメモリ部14、共有メモリ部15は、それぞれ独立したパッケージ(PK)、すなわち、チャネルIFPK1、ディスクIFPK2、

(14) 月2000-99281 (P2000-9BA)

セレクタPK3、メモリPK4にそれぞれ実装する。異なるキャッシュメモリ部14は、異なるパッケージ上に実装する。また、異なる共有メモリ部15も、異なるパッケージ上に実装する。1つのキャッシュメモリ部14と1つの共有メモリ部15は同一のパッケージ上に実装しても問題ない。図14では、1つのキャッシュメモリ部14と1つの共有メモリ部15を同一のメモリPK4上に実装した例を示している。

【0113】チャネルIFPK1、ディスクIFPK2、セレクタPK3、メモリPK4は、プラッタ5上に実装し、それを筐体に搭載する。

【0114】ここで、各PKをプラッタ5に実装するときには、セレクタPK3を中心付近より外側、好ましくはプラッタの両端に配置することが重要となる。このような配置により、各PK間を結ぶ線をプラッタ5上に配線する際、プラッタ5全体にわたって配線の密度を均一にでき、プラッタ上の配線を容易になる。

【0115】本実施例では、セレクタ部13をセレクタPK3に実装して、プラッタ5の両端に配置するとした。しかし、セレクタ部13をパッケージに実装せず、プラッタの両端に直接実装しても問題ない。

【0116】[実施例6]図1に示す実施例1のディスクアレイ制御装置1において、SMアクセス制御部105とSMコントローラ108間をアクセスバス2:137で接続する代わりに、図7に示すように、2本の共有バス130を介して接続する。SMアクセス制御部105からは、2本の共有バス130それぞれに1本ずつ接続バスを設ける。また、SMコントローラ108からも、2本の共有バス130それぞれに1本ずつ接続バスを設ける。SMアクセス制御部105及びSMコントローラ108は、それぞれ2つのアービタを有している。2つのアービタはそれぞれ2本の共有バスのアービトレーション用のアービタである。SMアクセス制御部105からSMコントローラ108へのアクセスの際は、複数のアービタの内の1つがマスタとなり、共有バス130の使用権のアービトレーションを行う。そして、使用権を得たSMアクセス制御部105がSMコントローラ108にアクセスを行う。また、アービタをSMアクセス制御部105及びSMコントローラ108内に設ける代わりに、独立した回路として共有バス130に直接接続しても問題ない。

【0117】共有バス接続では、バスのデータ幅を広げることによりデータ転送速度を上げることが可能であり、共有メモリ部15へのアクセス時間を短縮可能である。

【0118】上述したように、図1におけるアクセスバス2:137のデータ幅は、アクセスバス0:135のデータ幅よりも2倍以上小さくすることができるので、図1のようにチャネルIF部11及びディスクIF部12と共有メモリ部15との間をスター接続(1対1接

続)しても、共有メモリ部を実装するLSIのピンネックの問題が生じるケースは少ない。とはいっても、アクセスバス2:137の本数が増え過ぎてアクセスバス2:137を実装できないという問題が生じる可能性もある。そうした場合、本実施例の共有バス接続が有効となる。

【0119】なお、本実施例では、耐障害性の向上という観点から、セレクタ部13、キャッシュメモリ部14、及び共有メモリ部15をそれぞれ二重化しているが、これらを二重化しなくても、上述の効果を得られるることは言うまでもない。

【0120】

【発明の効果】本発明によれば、チャネルIF部、ディスクIF部—キャッシュメモリ間のアクセスバスについてはスループットを高くできる。また、チャネルIF部、ディスクIF部—共有メモリ間のアクセスバスについてはスループットを高くでき、かつアクセス時間を短くできる。これによって、スループットが高く、かつ応答時間の短いディスクアレイ制御装置を提供することができる。

【図面の簡単な説明】

【図1】本発明によるディスクアレイ制御装置の構成を示す図。

【図2】従来のディスクアレイ制御装置の構成を示す図。

【図3】従来のディスクアレイ制御装置の他の構成を示す図。

【図4】本発明によるディスクアレイ制御装置内のCMアクセス制御部の構成を示す図。

【図5】本発明によるディスクアレイ制御装置内のSMアクセス制御部の構成を示す図。

【図6】本発明によるディスクアレイ制御装置内のセレクタ部の構成を示す図。

【図7】本発明によるディスクアレイ制御装置内のキャッシュメモリ部の構成を示す図。

【図8】本発明によるディスクアレイ制御装置内の共有メモリ部の構成を示す図。

【図9】キャッシュメモリ部へのデータの書き込み時の手順を示す図。

【図10】共有メモリ部へのデータの書き込み時の手順を示す図。

【図11】二重化した2つのキャッシュメモリ部へデータを二重に書き込む時の手順を示す図。

【図12】1つのキャッシュメモリ部から別のキャッシュメモリ部へデータをコピーする時の手順を示す図。

【図13】二重化した2つの共有メモリ部へデータを二重に書き込む時の手順を示す図。

【図14】本発明によるディスクアレイ制御装置内の実装の構成を示す図。

【図15】本発明によるディスクアレイ制御装置内の他の構成を示す図。

(15) 月2000-99281 (P2000-9縫)

【図16】本発明によるディスクアレイ制御装置内の他の構成を示す図。

【図17】本発明によるディスクアレイ制御装置の構成を示す図。

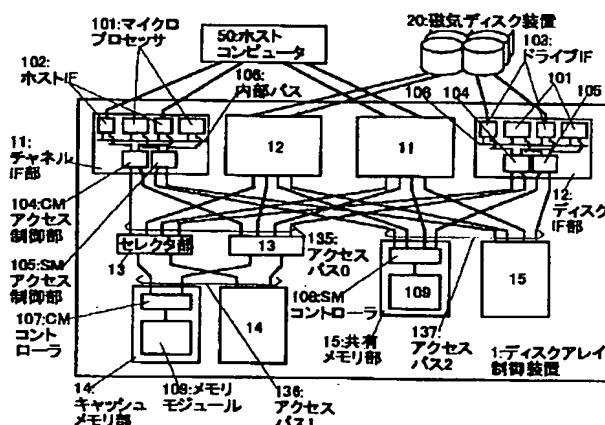
【符号の説明】

1…ディスクアレイ制御装置、11…チャネルIF部、12…ディスクIF部、13…セレクタ部、14…キャッシュメモリ部、15…共有メモリ部、20…磁気ディ

スク装置、50…ホストコンピュータ、101…マイクロプロセッサ、102…ホストIF、103…ドライブIF、104…CMアクセス制御部、105…SMアクセス制御部、106…内部バス、107…CMコントローラ、108…SMコントローラ、109…メモリモジュール、135…アクセスバス0、136…アクセスバス1、137…アクセスバス2。

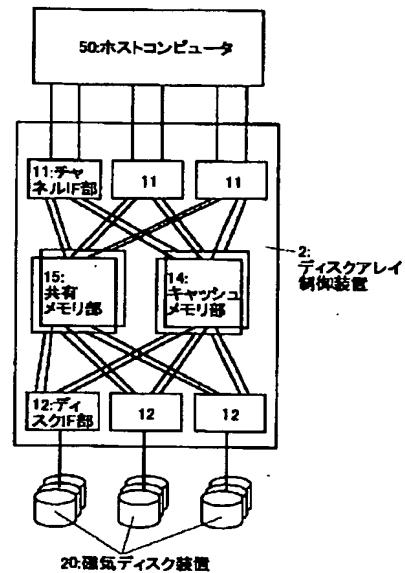
【図1】

図1



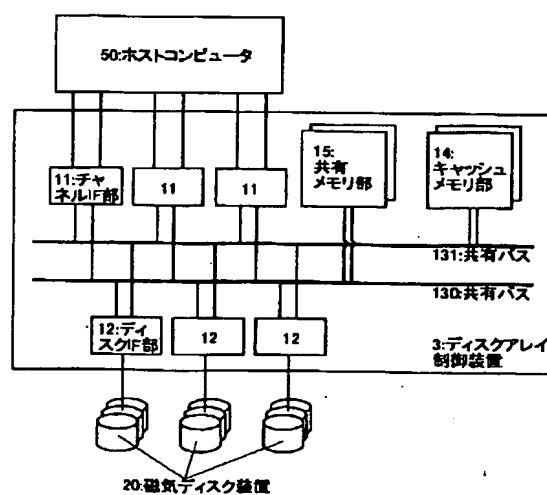
【図2】

図2



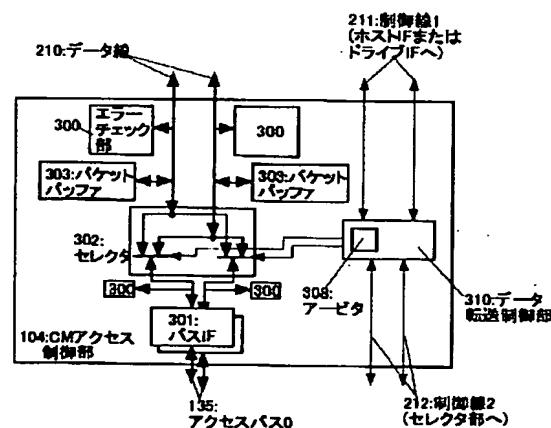
【図3】

図3



【図4】

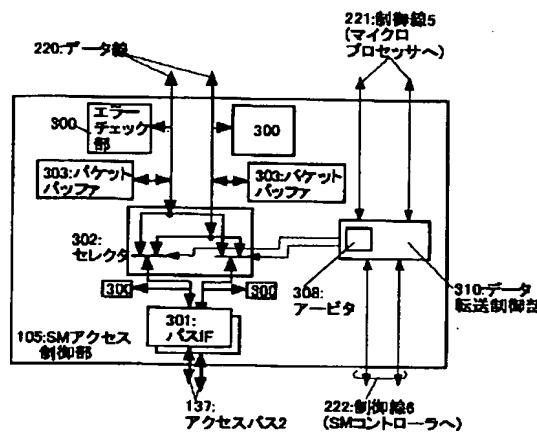
図4



(16)月2000-99281 (P2000-9\$ uA)

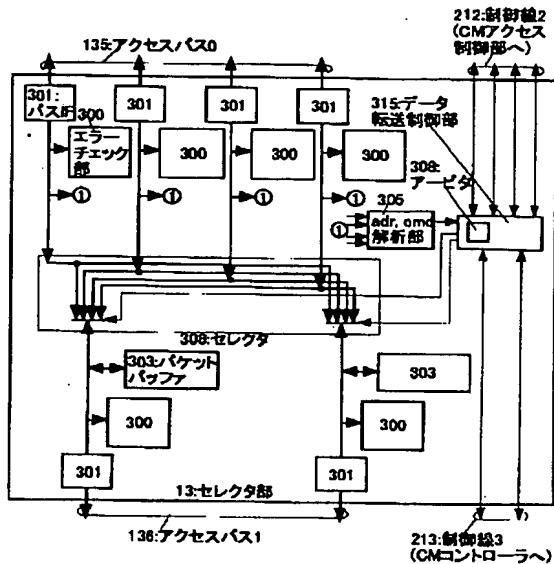
【図5】

図5



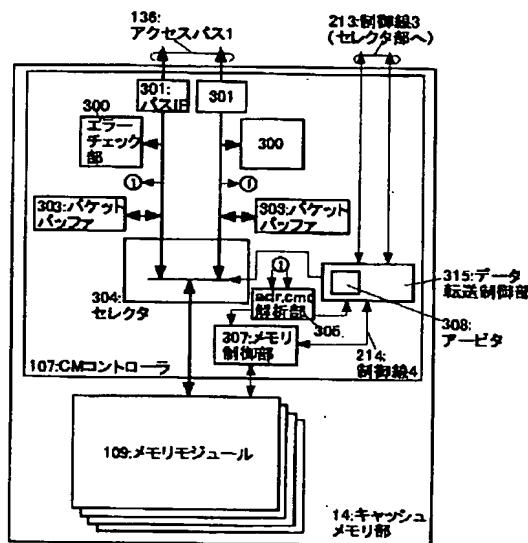
【図6】

図6



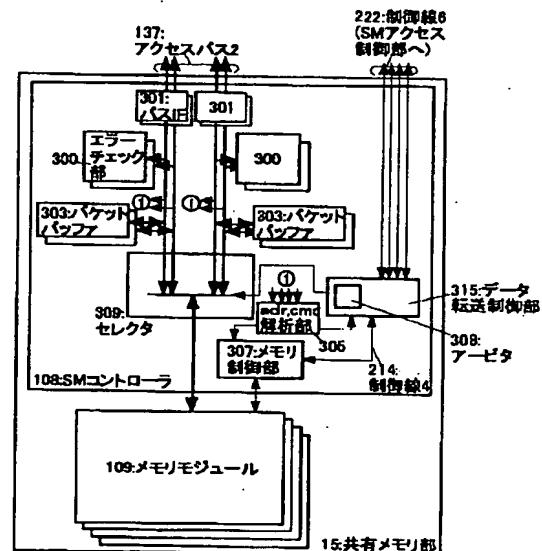
【図7】

図7



【図8】

図8



(17)月2000-99281 (P2000-94A)

【図9】

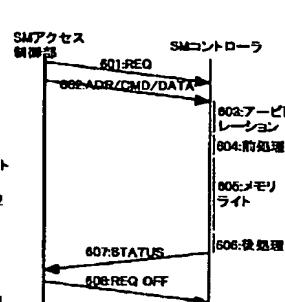
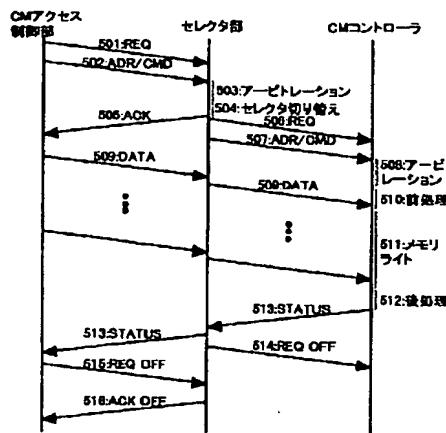
【図10】

【図14】

図9

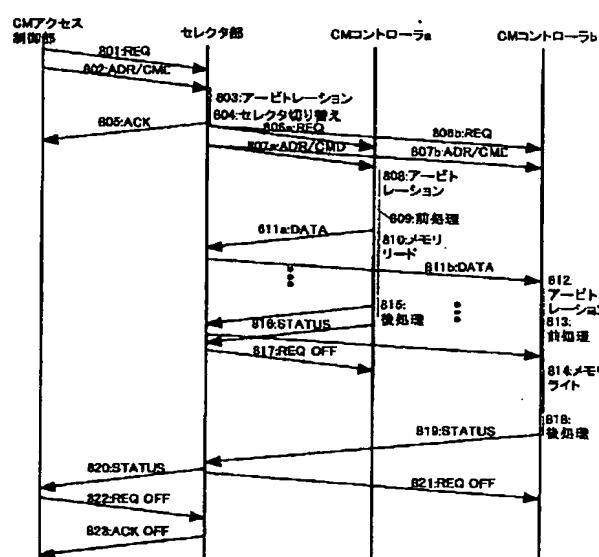
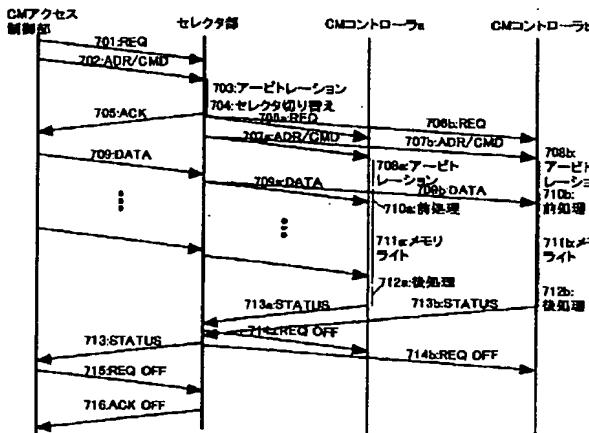
図10

図14



【図11】

図11



【図12】

図12

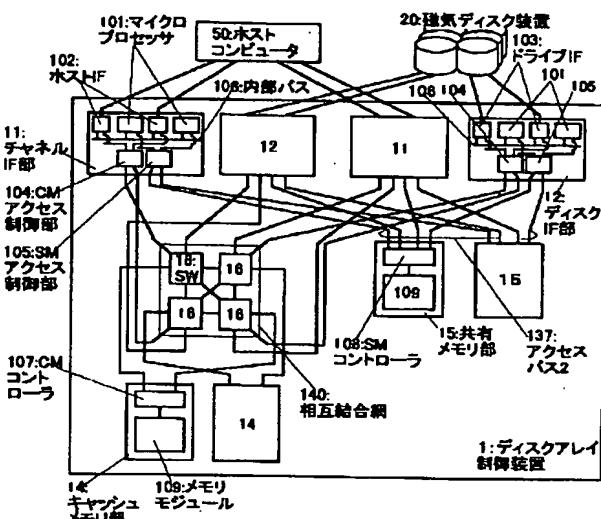
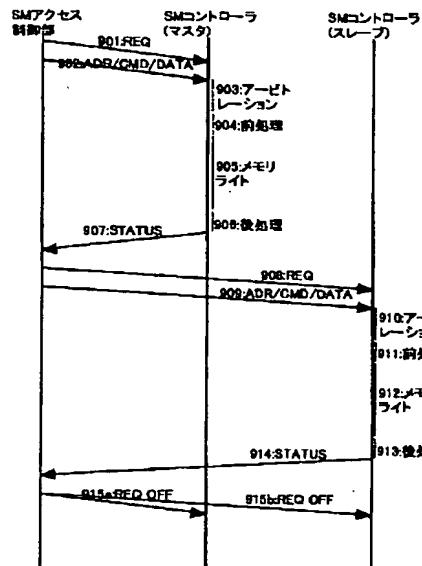
(18) 2000-99281 (P2000-9 磷織

〔図13〕

【図15】

圖 13

図15

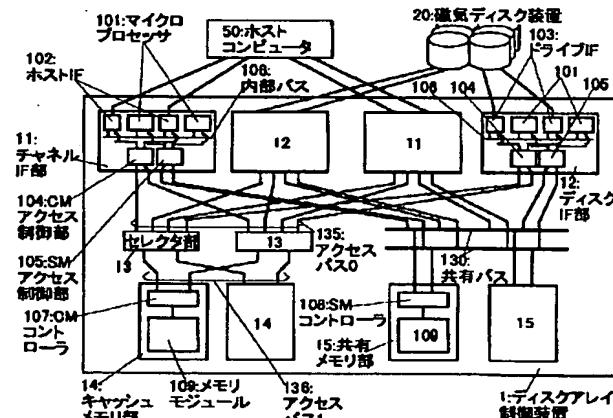
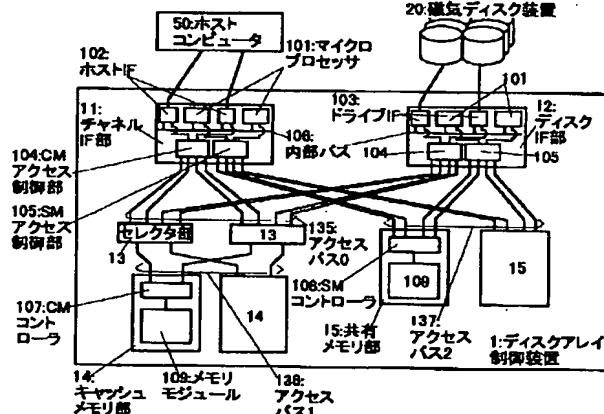


【図17】

【図16】

圖17

図16



フロントページの続き

(72) 発明者 藤林 昭
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(72) 発明者 金井 宏樹
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(19) 2000-99281 (P2000-9+A)

(72)発明者 箕輪 信幸

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

Fターム(参考) 5B065 BA01 CA12 CA30 CE11 CH01